

# Promoter Analysis *in silico*

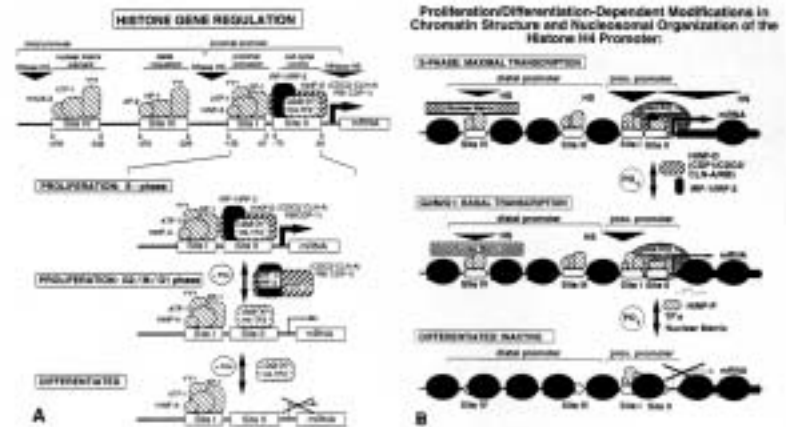
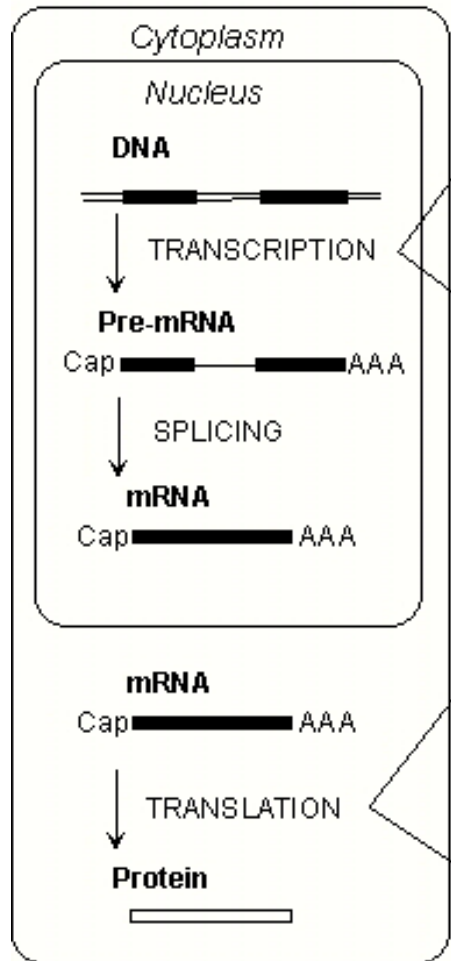
Michael Q. Zhang

Cold Spring Harbor Laboratory

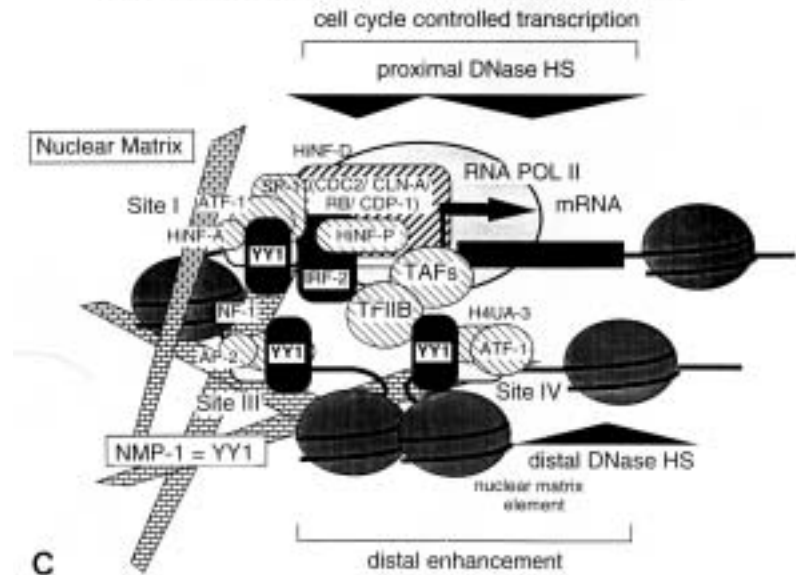
- Promoter: transcriptional control/regulation
- General promoter recognition problem
  - *De novo* prediction
  - Comparative analysis
- Specific promoter recognition problem
  - Database search and multiple alignment
  - Expression data analysis

# Promoter and transcription

## Eucaryotic cell

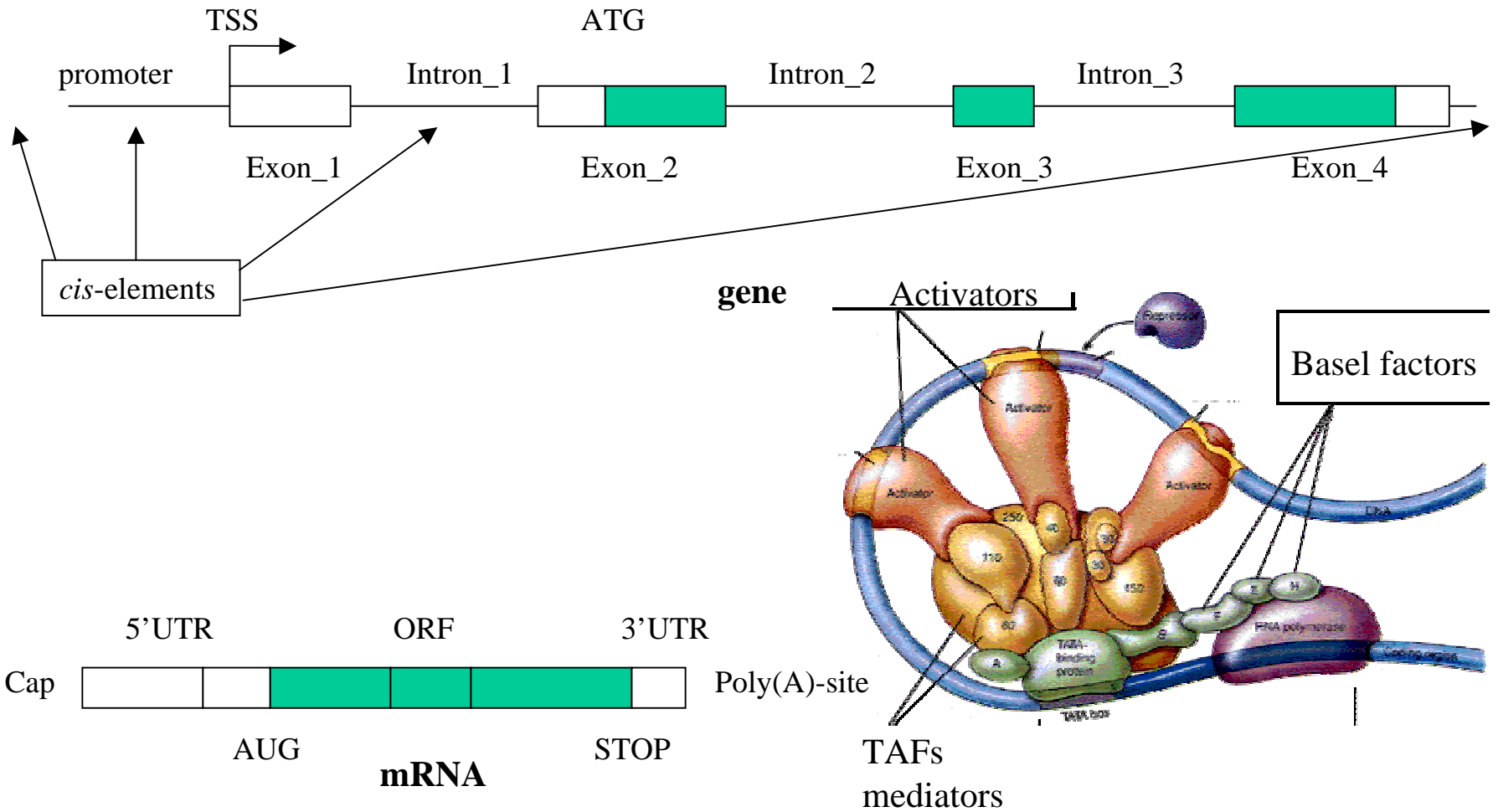


## Spatial Organization of the Histone Gene Promoter

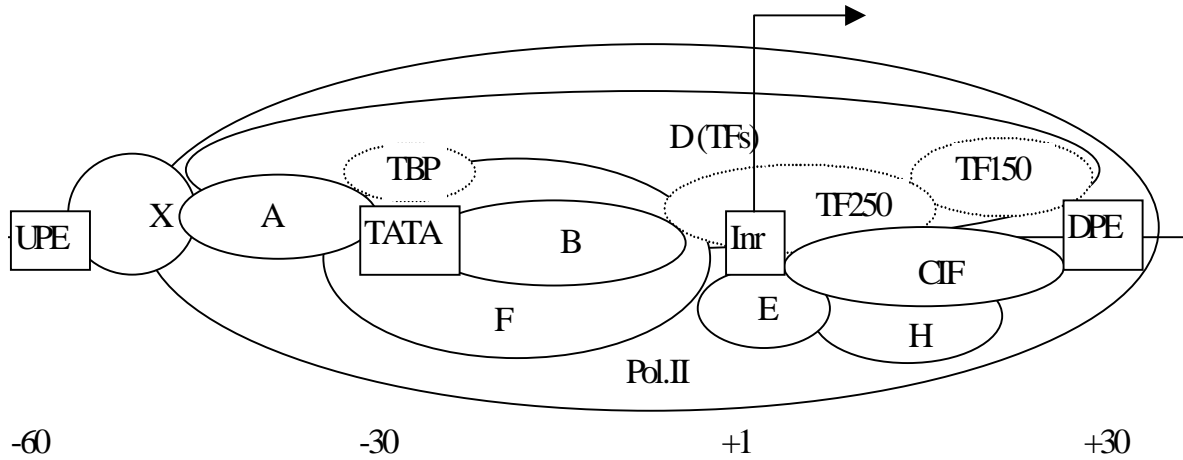


# General promoter recognition

## Gene structure prediction



# Core\_Promoter (Zhang, Genome Res. 1998)



GenBank

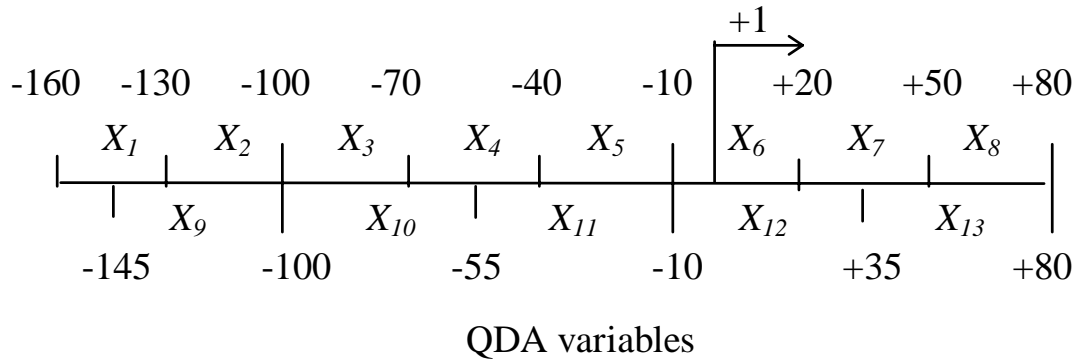
M12523:1..1980

ALB gene=serum albumin

Firstexon=1737..1854

AUG=1776

C+G=0.33



## Core\_Promoter prediction:

TSS	Score
<b>1737</b>	<b>0.637</b>
1736	0.604
1727	0.588
1732	0.534
1731	0.531
1728	0.498
1726	0.428

# CpG\_Promoter (Ioshikhes&Zhang *Nature Genet.* 2000)

CpG island: Length > 200 bp; C + G content > 50%; CpG ratio Obs/Exp > 0.6

- 135 genes
- 68 have CpG island around promoter
- 63 recognized
- SN = 0.47 (0.93)
- SP = 0.34 (1 Pos./26 kb; 1/36 kb is in fact)

•Promoter Scan gives

SN = 0.44

SP = 0.06 (1 Pos. / 4.7 kb)

GenBank	CpG_Promoter prediction:			Core_Promoter prediction:	
	CpG islands	Promoter-associated		TSS	Score
D87675	<b>8813..9319</b>	+		<b>8921</b>	<b>0.100</b>
>301kb	<b>9328..9547</b>	+		8923	0.094
App gene encodes	<b>9761..10203</b>	+		8920	0.089
Amyloid precursor protein	117256..117511	-		8919	0.084
Firstexon=9001..9204	176132..176342	-		8922	0.078
AUG=9148	257735..257942	-		8918	0.058
	261475..261750	-		8783	0.056

# First exon prediction (FirstEF)

(Daluvuri,Grosse&Zhang, submitted)

## Performance statistics of FirstExonFinder based on cross validation

Exon Type	Sn	Sp	CC
CpG-related	0.92	0.97	0.94
Not CpG-related	0.74	0.6	0.65
All Exons	0.86	0.83	0.83

## Promoter Prediction accuracy of FirstEF and PromoterInspector for Ch22

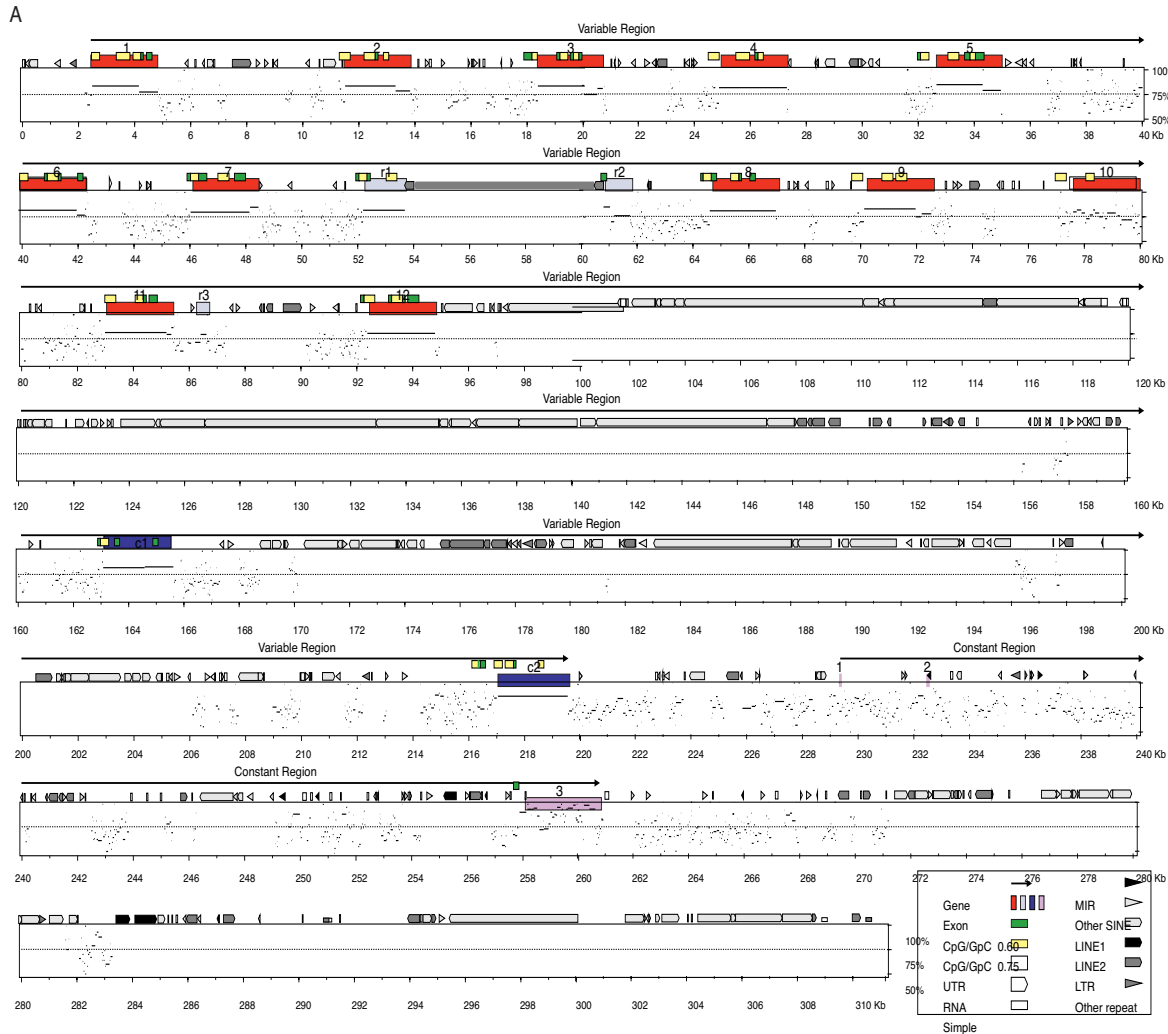
Program	TP	FP	Sn	Sp
FirstEF	46	40	79.30%	53.50%
PromoterInspector	28	37	48.30%	43.10%

## Prediction Accuracy for Ch21&22 (Number of Real Promoters: 58)

Chromosome	Number of Exponentially mapped first exons	Number of correctly predicted first exons	Completely non-coding exons	Predicted non-coding exons
21	42	37 (88%)	14	10 (71%)
22	79	69 (87%)	28	23 (82%)
Total	121	106 (88%)	42	33 (79%)

# Comparative genomics (Wu et al. Genome Res. 2001)

Maniatis lab (Harvard), Zhang lab (CSHL) & Myers lab (Stanford)



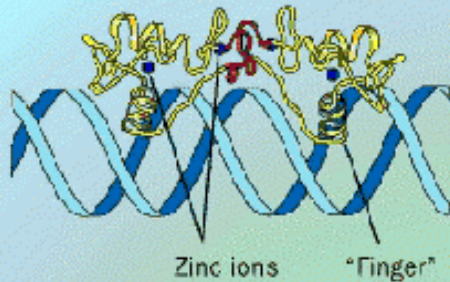
mPcdhy-a1	-260	ga	CTCTGGCGCC	cgct	GTTCAACC	aa	-240	(1.00)
mPcdhy-a2	-262	gc	CTCTGAGCGT	cgct	GTTGACC	ac	-242	(1.00)
mPcdhy-a3	-234	gc	CTCTGGGTGT	cgct	GTTGACT	ac	-214	(0.90)
mPcdhy-b1	-257	ac	CATTAGCGGA	cgct	GTTGACC	aa	-237	(0.60)
mPcdhy-a4	-231	tg	CTCTGGGTGT	cgct	GTTCAACC	ac	-211	(1.00)
mPcdhy-b2	-206	ac	CTCTGAGCGT	cgct	GTTAGGTC	aa	-186	(0.80)
mPcdhy-a5	-195	ga	CTCTAAGCGT	cgct	GTTCAACC	ag	-175	(1.00)
mPcdhy-a6	-239	ga	CTCTAAGTGT	cgct	GTTCCCT	aa	-219	(0.90)
mPcdhy-a7	-244	ga	CTCTGCCGCC	cgct	GTTCAACC	aa	-224	(1.00)
mPcdhy-b4	-242	ga	CTCTGGCGCC	cgct	GTTGGCC	ag	-222	(1.00)
mPcdhy-a8	-236	ga	CTCTAAGCGT	cgct	GTTCAACC	ta	-216	(1.00)
mPcdhy-b5	-242	aa	CTCTGGCGCC	cgct	GTTGGCC	tg	-222	(1.00)
mPcdhy-a9	-256	ga	CTCTGAGCGC	cgct	GTTCAACC	aa	-236	(1.00)
mPcdhy-b6	-232	gc	CTCTGGCGCC	cgct	GTCGGCC	ag	-212	(0.86)
mPcdhy-a10	-256	ga	CTCTGAGCGC	cgct	GTTCAACC	ta	-236	(1.00)
mPcdhy-b7	-250	gc	CTCTGGCGCC	cgct	GTCGGCC	ag	-230	(1.00)
mPcdhy-a11	-258	gg	CTCTAGCGCC	cgct	GTTGACC	aa	-238	(1.00)
mPcdhy-b8	-225	gc	CTCTGGCGCC	cgct	GCTGACC	aa	-205	(0.86)
mPcdhy-a12	-222	ga	CTCTGAGCGC	cgct	GTTCAACC	ac	-202	(1.00)
mPcdhy-c3	-230	ga	CTCTGTGTGC	cgct	GTCGGCC	aa	-210	(1.00)
mPcdhy-c4	-206	ga	GCCAGAGATT	act	GCTGACC	aa	-186	(0.12)
mPcdhy-c5	-239	tg	CTCTTCCTGC	TGT	CTCTCTCC	cc	-219	(0.10)
Consensus			CTCT	G G	cgct	GT	CC	
hPcdhy-a1	-275	ga	CTGTGGCGCC	cgct	GTTCAACT	aa	-255	(0.71)
hPcdhy-a2	-262	gc	CTCTGAGCGT	cgct	GTTGACC	ac	-242	(1.00)
hPcdhy-a3	-260	gc	CTCTTAGTGT	cgct	GTTGACC	ac	-240	(1.00)
hPcdhy-b1	-229	ac	CTCTGGGCGC	cgct	GTCGACC	aa	-209	(1.00)
hPcdhy-a4	-234	ga	CTCTGGGTGT	cgct	GTTCAACC	aa	-214	(1.00)
hPcdhy-b2	-208	ac	CTCTGAGCGT	cgct	GTTGGTC	aa	-188	(0.86)
hPcdhy-a5	-217	ga	CTCTGAGTGT	cgct	GTTGACC	aa	-197	(1.00)
hPcdhy-b3	-226	gg	CTCTGGGCGC	cgct	GTTGGTC	ag	-206	(0.42)
hPcdhy-a6	-242	ga	CTCTAAGCGT	cgct	GTTGACC	aa	-222	(1.00)
hPcdhy-a7	-244	ga	CTCACAGCGC	cgct	GTTCAACC	aa	-224	(1.00)
hPcdhy-b4	-261	ga	CTCTGGGCGC	cgct	GTTGACC	aa	-241	(1.00)
hPcdhy-a8	-239	ga	CTCTAAGCGC	cgct	GTTCAACC	ta	-219	(0.85)
hPcdhy-b5	-258	ga	CTCTGGGCGC	cgct	GTTGGCC	aa	-238	(1.00)
hPcdhy-a9	-232	ga	CTCTGAGCGC	cgct	GTTGACT	aa	-212	(0.86)
hPcdhy-b6	-272	gc	CTCTGGGCGC	cgct	GTCGGCC	ag	-252	(1.00)
hPcdhy-a10	-238	ga	CTCTGAGCGC	cgct	GTTCAACC	ta	-218	(1.00)
hPcdhy-b7	-270	gc	CTCTGGGCGC	cgct	GTCGGCC	ag	-250	(0.78)
hPcdhy-a11	-249	gg	CTCTAGGCGC	cgct	GCTCACC	aa	-229	(0.84)
hPcdhy-a12	-227	ga	CTCTGAGCGC	cgct	GTTCAACC	aa	-207	(1.00)
hPcdhy-c3	-230	ga	CTCTGTGTGC	cgct	GTCGGCC	aa	-210	(1.00)
hPcdhy-c4	-215	ga	GCCGGGAG	cgct	GCTGACC	aa	-195	(0.57)
hPcdhy-c5	-186	ag	CTCCCTCCC	act	GCCACC	cc	-166	(0.20)
Consensus			CTCT	G G	cgct	GT	CC	

Pipmaker

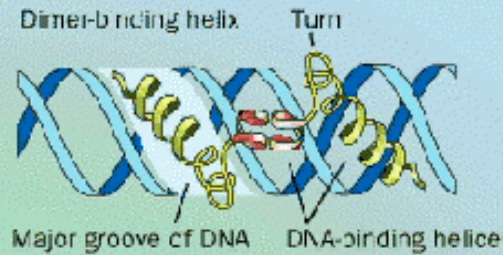
Gibbs sampler

# Specific promoter recognition (Gene expression prediction)

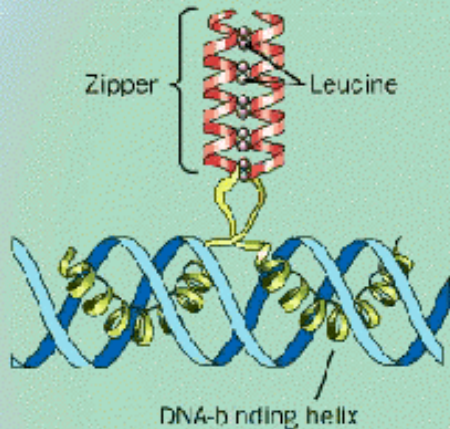
**Zinc finger**



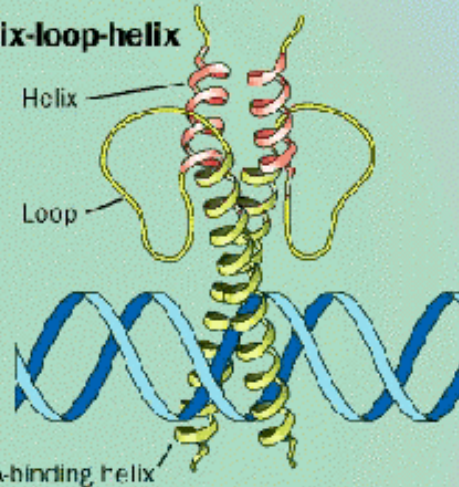
**Helix-turn-helix**



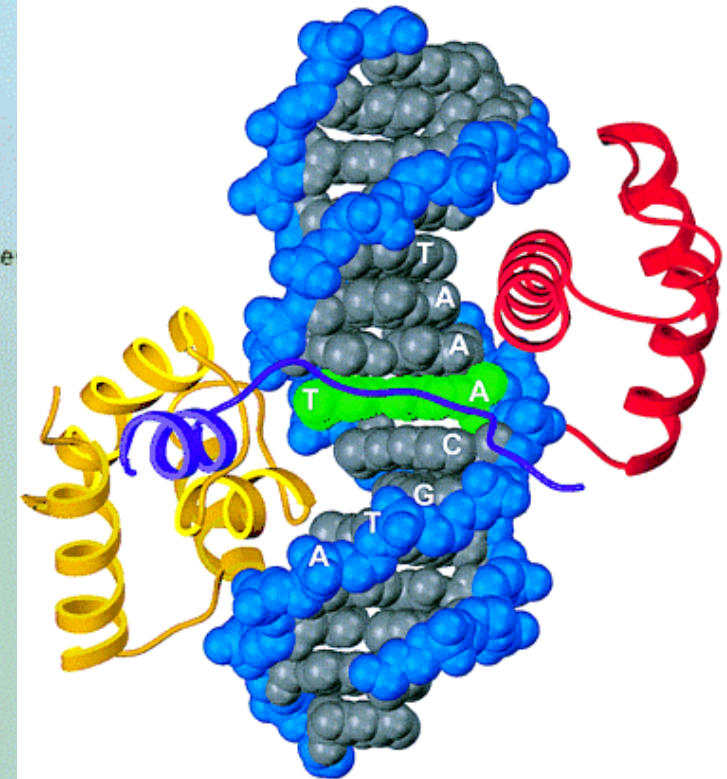
**Leucine zipper**



**Helix-loop-helix**



**A**





# Liver-specific Promoter Database

(LSPD: Zhang&Zhang unpublished)



LSPD

*The Liver Specific Gene Promoter Database*



LSPD

Genes

Sites

Composite Sites

Factors

Tools

Links

## Liver Specific Genes

- List ~300 promoter regions responsible for liver specific transcriptions
- Collect ~400 experimentally verified regulatory regions and elements
- Provide information on transcription regulation of liver genes
- Compare transcription regulation of *functionally* or *evolutionarily* related genes
- Retrieve sequences of the promoter region

## Regulatory Elements

- Provide information on transcription regulatory elements
- Report the methods for verification of the elements
- Record binding affinity and regulatory function
- Summarize the site distribution and sequence consensus

## Composite Sites

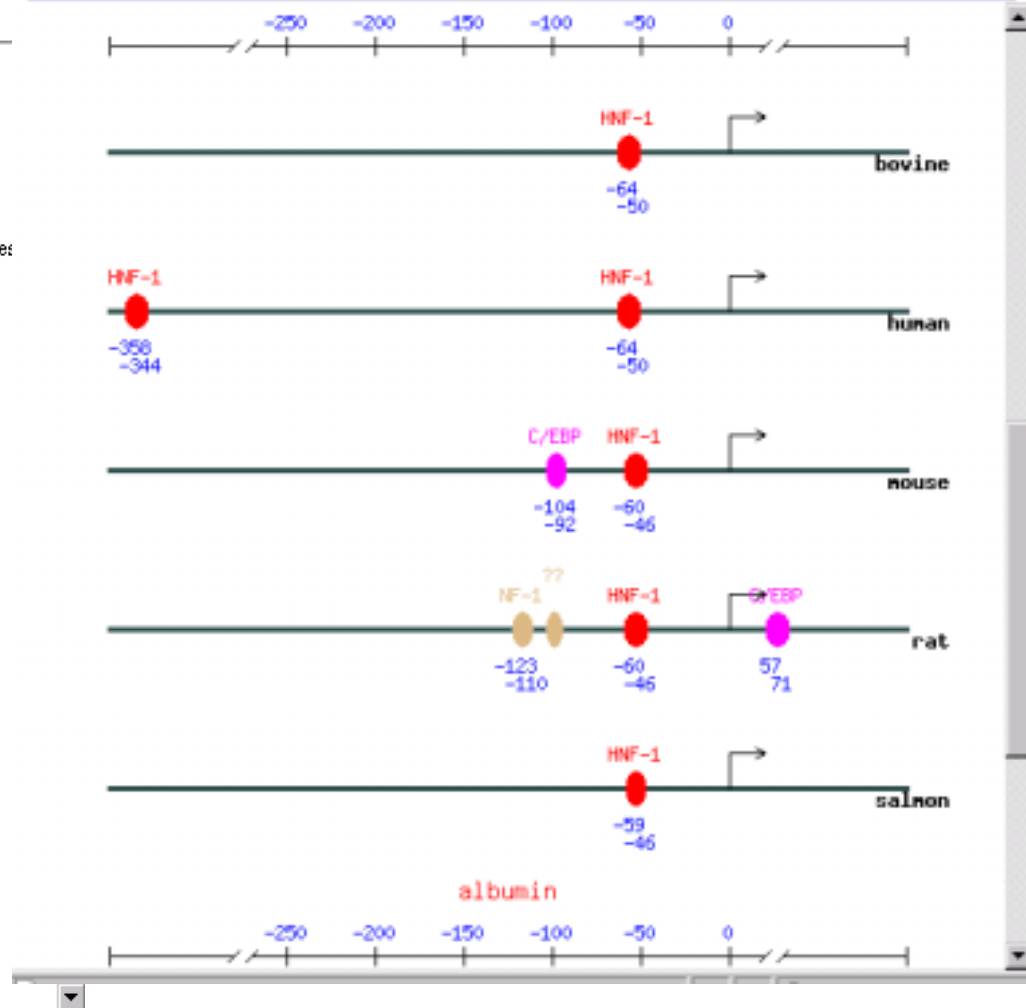
- Record interaction among transcription factors
- Explore space constraints and correlation of regulatory elements
- Characterize the condition for factors to interact

## Transcription Factors

- Provide alignment of all experimentally verified sites
- Construct matrix and consensus sequence
- Summarize information on the factors and their roles in regulation

## Analysis tools

- Retrieve promoter sequences
- Discover new motifs
- Search known motifs
- Build promoter models



# LSPD (continued)



- LSPD
- Genes
- Sites
- Composite Sites
- Factors
- Tools
- Links

## albumin

<b>Brief Name</b>	albumin
<b>Function</b>	(Transport/Binding)
<b>Cell</b>	liver

### Regulatory regions of ortholog genes:

	Accession #	TSS	ATG	# known sites	include?
<a href="#">bovine</a>	<a href="#">X62448</a>	241		1	<input type="checkbox"/>
<a href="#">human</a>	<a href="#">M92816</a>	enhancer		1	<input type="checkbox"/>
<a href="#">human</a>	<a href="#">M12523</a>	1737		2	<input type="checkbox"/>
<a href="#">mouse</a>	<a href="#">M63182</a>	enhancer		5	<input type="checkbox"/>
<a href="#">mouse</a>	<a href="#">J04738</a>	2043		2	<input type="checkbox"/>
<a href="#">rat</a>	<a href="#">M16825</a>	153		4	<input type="checkbox"/>
<a href="#">salmon</a>	<a href="#">X79487</a>	299		1	<input type="checkbox"/>
<a href="#">xenopus laevis</a>	<a href="#">Z26826</a>	1518		1	<input type="checkbox"/>
<a href="#">xenopus laevis</a>	<a href="#">Z26825</a>	1607		1	<input type="checkbox"/>

Retrieve regulatory sequences from  to  with respect to

Multiple sequence alignment

### CLUSTAL W (1.8) multiple sequence alignment

Gene [albumin](#)

```

X.laevis_Z26825      CACACGAAATACGCTTATTTGACCTTAAAGTGTGATTC----ACATTAGZAAATTCACA
X.laevis_Z26826      CACACGAAATACATTTATTTGACCTTAAAGCTTATTTG----ACITTAGZAA-CTCCACA
Mus_J04738          -ACATACGCAAGGZATTTAGTCAAAACAACITTTTGCZAAAGATGTATCA--ITTTGIA
Rat_M16825          -GATTTAGTAAACAACITTTTTT--ITTCITTTTGCZAAAGATGTATCA--ITTTGIA
E.tSueus_X62448     -ACATACACAGGZATTTAGTCAAAACAATTTTTTGCZAAAGATGTATGAA--ITTTGIA
Human_M12523        -AATATACACAGGZATTTAGTCAAAACAATTTTTTGCZAAAGATATATGAA--ITTTGIA
S.gallus_X79487     ACACACAGCAGACAGCAGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG
          *          *          *          *          *          *          *          *          *          *
X.laevis_Z26825      AAGCTAAAACCACTGC--AAACACACAAATTTGATAGGTAAATAAATTTCCAGATCTCTC
X.laevis_Z26826      AAGCTAAAACCACTGC--AAACACACATTTTGTAGGTAAATCAATTTTCCAGATCTCTC
Mus_J04738          ATGGGTAGACCAACCAATGAAATGCGAGG--TAGTATGCTTAAATGATCTACAGTATATGGT
Rat_M16825          ATGGGTAGACCAACCAATGAAATGCGAGG--TAGTATGCTTAAATGATCTACAGTATATGGT
E.tSueus_X62448     ATCACTTATGACCAATGAAATACAAAGATGAGTCAGTAAATGATCTACAAATATATGGC
Human_M12523        ATCGTTCGACCAATGAAATACAAAGATGAGTCAGTAAATGATCTACAAATATATGGT
S.gallus_X79487     AGACAGACAGCAGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG
          *          *          *          *          *          *          *          *          *          *

```

### HNF-1

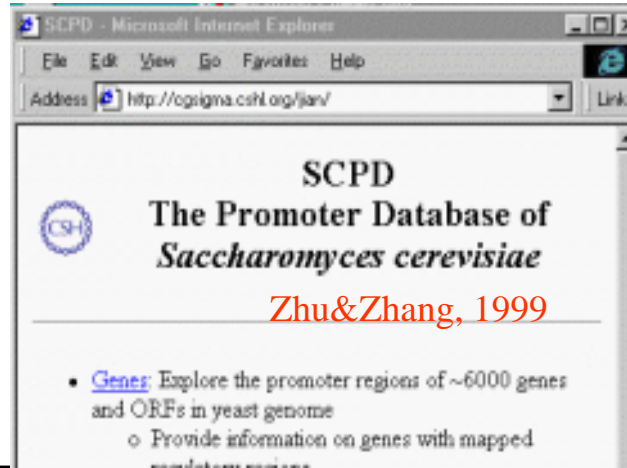
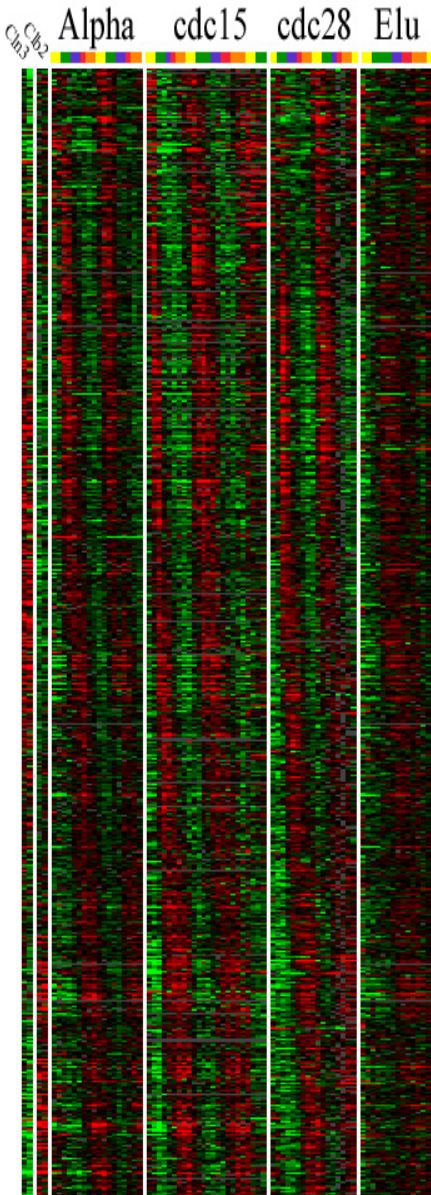
TFs	#exp. sites	motif
<i>HNF1</i>	47	(A/G)TTAAT(h) (IR(1))
<i>HNF3</i>	54	TRTT(G/T)RY (core)
<i>HNF4</i>	37	AGGTCA(h) (DR(1(2)))
<i>C/EBP</i>	80	GCAAT(h) (palindrome)

### •C/EBP frequency matrix

pos	1	2	3	4	5	6	7	8	9	10
A	29	6	1	21	18	3	15	68	68	13
C	16	6	3	61	7	10	46	10	6	16
G	20	6	20	34	20	58	0	2	0	14
T	15	62	56	19	25	9	19	0	6	37

# Cell Cycle Regulation (Spellman *et al.* MBC,1998)

(Brown&Botstein labs – Futcher&Zhang labs)



SWI5 TAACCTGTTTAGGAAAAAGGTAACAATAAC  
 CLB1 CCGCCCAAAGAGGAAAAATCATCAACAATCAA  
 CLB2-1 CGACCGAATCAGGAAAAAGTCAACAACGAAG  
 CLB2-2 TTTCCCTAAA CGGCTCAAATATGTAACAATAT  
 BUD4 TGACCCGATTTGGAAAAAGGTAACAACAAT  
 KIN3 TTTCCCTAATTAGGTTAAACGTAAATAAAGC  
 CDC20 TTGGCCGAAAGAGGCAAAACGTAAATAGGTTG  
 YOR315W TCGCCCAATTAGGATAAAAGTAAATAACATA  
 CHS2 ATACTCAAATAGGAAATATATAACAATAAG  
 YRO2 CAACCCGATGAGGAATCATCCCGATCTAACCA  
 WSC4 TCGCCCAAAGTGGAATAACCGTAATAGTACC  
 APC11 TGACCGGTAGTGGAACTGTTCCAAACCTTTT  
 FAR1 ATACCTAAAAGGAAATAGGCAATAAAATGA  
 CDC5 AAACCCAATAAGAAAAATCCAAATATAGAA  
 YML119W TTTCCAGATTAGGAAAGACATAAATAAAAA  
 UBA2 TTTTCTGTATTGGTAAGAACTTAAATGAGAA  
 DBF2 TTTTCTTTTGGGTTGGTCTCCCGAATAGG  
 CDC16 TTTCTTTATTGGGTCACAAACCGTATTAA

Name	Genes	Group	Motifs	Sites (%Genes)	Sites (%Genes out of 256 controls)	Cho et al.
Cln2	58	G1	MCB:ACGCGT SCB:CRCGAAA	52(62) 43(52)	15(6) 33(13)	+ +
Y'	31	G1	RAP1:TGCACCW ?:AGCSGCT, etc.	42(71) 32(52)	33(12) 16(3)	- -
Fks1	38	G1	SCB:CRCGAAA ?:TKCAKCTGCA	26(53) 4(11)	33(13) 3(1)	+ -
Histone	9	S	CCA:GcGAARYtngRGAACr NEG:CATTgnGCG SCB:CGCGAAA	19(100) 18(89) 7(56)	0(0) 1(0) 14(5)	- - +
Met	20	S/G2	Cbf1/Met/Met28:TCACGTG Met31/Met32:AAAnTGTGG	20(60) 14(55)	17(5) 12(5)	- -
Clb2	36	G2/M	Mcm1(P-box):TTWCCYrawnnGGaAA Mcm1(P-box)+Sff:(P)n <sub>2-4</sub> GTaAAYAA	55(64) 19(47)	1(0) 0(0)	+ -
Mcm	38	M/G1	ECB:TTTCCcaATngGGAAA ?:AAAGAAAA	73(79) 26(53)	1(0) 20(8)	+ -
Sic1	27	M/G1	SWI5:RRCCAGCR ?:GCSCRGC	23(48) 12(41)	23(9) 31(11)	- -
Mat	13	M/G1	Ste12(PRE):TGAAACA P'+Q:tTTCCTaaTTrGknnnTCAATG ?:WnAnnAGCCAnnnnWwNMAAAAnA	10(54) 8(46) 6(46)	48(18) 0(0) 2(1)	- - -

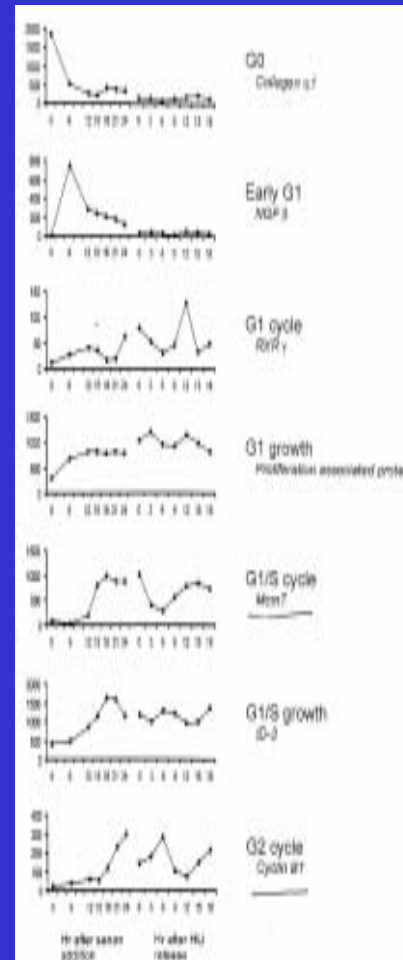
# Cell Cycle Regulation (continued)

Computationally predicted E2F target genes confirmed by *in vivo* footprint (ChIP) (Kel *et al.* JMB 2001)

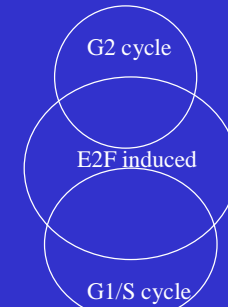
Gene	EMBL	Sequence of the potential sites	Position rel. start transcription	Score, q	d(X)
<b>c-fos</b> , <i>Homo sapiens</i>	HSFOS	(-) gcCTTGGCGCGTGTcc	-165 .. -176	0.915	2.92
		(-) ggGGTGGCGCGCGGgc	-92 .. -103	0.836	
		(+) ccTCTGGCGCCACCgt	-90 .. -79	0.878	
		(-) acGGTGGCGCCAGAgg	-78 .. -89	0.830	
<b>JunB</b> , <i>Homo sapiens</i>	HS207341	(+) gcTATGCGCCAGAgg	79 .. 90	0.887	3.16
		(-) tcTCTGGCGCGATAgc	91 .. 80	0.905	
		(-) ggGCTGGCGCGGGCgg	169 .. 158	0.820	
<b>TGF-β 1</b> , <i>Homo sapiens</i>	HSTGFBIPR	(+) ctGTTTGGCGGGCGga	-513 .. -502	0.804	2.03
		(+) ccCTTGGCGCCCTGgg	-298 .. -287	0.912	
		(+) ctCTTGGCGCGACGct	28 .. 39	0.928	
		(-) agCGTCGCGCCAAAgg	40 .. 29	0.830	
		(+) ccTTTGGCGCGGGga	85 .. 96	0.854	
<b>ARF</b> , <i>Homo sapiens</i>	AF082338	(-) acTTTCCCGCCCTGtg	-265 .. -276	0.859	
<b>Mcm4</b> (Cdc21), <i>Mus musculus</i>	AB000629	(+) ggTTTCCCGCCAAAac	-443 .. -432	0.872	
		(-) gtTTTGGCGGGAAAcc	-431 .. -442	0.935	
		(+) gcAGTGGCGCCTTCcg	-329 .. -318	0.810	
		(+) ccTTTGGCGCGTGTgat	-297 .. -286	0.846	
		(+) tgGGTGGCGCGAGAAct	-127 .. -116	0.809	
		(+) ttTGTCGCGCAGCAac	-24 .. -13	0.858	
<b>MCM5</b> (P1-CDC46), <i>Homo sapiens</i>	HS286B10	(+) agTTTGGCGCCAAAAtt	-187 .. -176	0.988	4.91
		(-) aaTTTGGCGCGAAAAct	-175 .. -186	1.005	
		(+) ttTTTCCCGCGAAAAct	8 .. 19	0.885	
		(-) agTTTCCCGGGAAAaaa	20 .. 9	0.932	
<b>von Hippel-Lindau</b> (VHL), <i>Homo sapiens</i>	AF010238	(+) aagcTTCGCGCCAGTgc	-270 .. -259	0.810	2.22
		(-) gcAGTGGCGCGAGctt	-258 .. -269	0.838	
		(-) gtCTTGGCGCGCGctc	-28 .. -39	0.921	
<b>B-myb</b> , <i>Homo sapiens</i>	HSBMYBDNA	(-) gtCCTGGCGCGCGGgc	-72 .. -83	0.831	5.50
		(-) gcCTGGCGCGCGGgc	-53 .. -42	0.866	
<b>Nucleolin</b> , <i>Homo sapiens</i>	HSNUCLEO	(-) cCTTGGCGCGCGGctg	-297 .. -308	0.966	2.91
		(-) ccGTGGCGCGCGGgt	-256 .. -267	0.814	
<b>Nucleolin</b> , <i>Cricetus griseus</i>	CSNUCLEO	(-) cgTTTGGCGCGCGTtg	-296 .. -307	0.973	6.67
<b>Nucleolin</b> , <i>Mus musculus</i>	MMNUCLEO	(-) agTTTGGCGCGCGctg	-306 .. -317	0.973	1.76

## Mammalian cell cycle regulation

(Ishida *et al.* MCB, 2001) (Nevins, Zhang labs, unpublished)



CDE.....CHR  
 Cdc2 agcgcgggtgagtttgaact  
 cdc25C ctggcgggaaggttgaatgg  
 Cyc-A2 gtcgcgggatacttgaactg  
 Cyc-B1 gcagtgcggggtttaaactc  
 Cyc-B2 gcggcgcggtatttgaatcg  
 Ki-67 ggcgcgccgggctggaagaa  
 IHABP tgggcggcgg-taggaatct  
 H2A.X ccagaggtgggtattaaaaa  
 ...



E2F  
 Cyc-E gTTCGCGCGCAGgg  
 P107 tTTTCGCGCGCTTtg  
 B-myb aCTTGGCGGGAGAta  
 Mcm4 gTTTCCCGCCAAAac  
 Mcm5 gTTTCGCGCCAAAat  
 tTTTCCCGCGAAAAct  
 DNAPola aCAGGGCGCCAAAacg  
 Rad51 gTTTGGCGGGAATtc  
 ...

# Sample Web Resources

## Promoter finding

- [CSHL](http://www.cshl.org/mzhanglab): [www.cshl.org/mzhanglab](http://www.cshl.org/mzhanglab)
- PromoterInspector: [www.gsf.de/biodv](http://www.gsf.de/biodv)
- PromoterScan: [bimas.dcert.nih.gov/molbio](http://bimas.dcert.nih.gov/molbio)
- Promoter2.0: [www.cbs.dtu.dk/services/promoter](http://www.cbs.dtu.dk/services/promoter)
- NNPP: [www-hgc.lbl.gov/projects/promoter.html](http://www-hgc.lbl.gov/projects/promoter.html)
- TSSG/TSSW: [dot.imgen.bcm.tmc.edu:9331/gene-finder/gf.html](http://dot.imgen.bcm.tmc.edu:9331/gene-finder/gf.html)

## Motif finding

- Gibbs sampler: [bayesweb.wadsworth.org/gibbs/gibbs.html](http://bayesweb.wadsworth.org/gibbs/gibbs.html)
- MEME: [meme.sdsc.edu/meme/website/](http://meme.sdsc.edu/meme/website/)
- Consensus: [bioweb.pasteur.fr/seqanal/interfaces/consensus-simple.html](http://bioweb.pasteur.fr/seqanal/interfaces/consensus-simple.html)
- ClustalW: [www.ebi.ac.uk/clustalw/](http://www.ebi.ac.uk/clustalw/)

## Databases

- TRANSFAC: [transfac.gbf.de/TRANSFAC](http://transfac.gbf.de/TRANSFAC)
- TRRD/COMPEL: [compel.bionet.nsc.ru/](http://compel.bionet.nsc.ru/)
- EPD: [www.epd.isb-sib.ch](http://www.epd.isb-sib.ch)
- MTIR: [www.cbil.upenn.edu/MTIR/HomePage.html](http://www.cbil.upenn.edu/MTIR/HomePage.html)
- LSPD (CSHL)
- SCPD +Fly+Worm+Plant (CSHL)

## Comparative genomics tools

- PIPMaker: [bio.cse.psu.edu/pipmaker](http://bio.cse.psu.edu/pipmaker)
- VISTA: [www-gsd.lbl.gov/vistal](http://www-gsd.lbl.gov/vistal)



# Zhang Lab Members 2001

